Key Reports to Evaluate z16 Processor Caches

Peter Enrico

Email: Peter.Enrico@EPStrategies.com

z/OS Performance
Education, Software, and
Managed Service Providers

Creators of Pivotor®

Enterprise Performance Strategies, Inc.
3457-53rd Avenue West, #145
Bradenton, FL 34210
http://www.epstrategies.com
http://www.pivotor.com

Voice: 813-435-2297
Mobile: 941-685-6789

Contact, Copyright, and Trademark Notices

Questions?

Send email to Peter at Peter.Enrico@EPStrategies.com, or visit our website at http://www.epstrategies.com or http://www.pivotor.com.

Copyright Notice:

Trademarks:

Enterprise Performance Strategies, Inc. presentation materials contain trademarks and registered trademarks of several companies.

The following are trademarks of Enterprise Performance Strategies, Inc.: **Health Check®, Reductions®, Pivotor®**

Other trademarks and registered trademarks may exist in this presentation

Abstract

- **Key Reports to Evaluate z16 Processor Caches**

  - This presentation will walk through, and explain, several reports that will be useful when evaluating the primary processor cache measurements on the z16 processor. There will be a review of the basic concepts and usage of the processor caches and then show which reports and measurements should be used to assess the effects of processor caches in the z16 environment.

EPS: We do z/OS performance…

- Performance reporting and analysis of your SMF data
  - Pivotor: Not just reporting, but analysis-based reporting based on our expertise

- Performance Educational Workshops (while analyzing your own data)
  (All presented via the web)
  - Essential z/OS Performance Tuning
    - March 20 – 24, 2023
  - Parallel Sysplex and z/OS Performance Tuning
    - February 7 – 8, 2023
  - WLM Performance and Re-evaluating Goals
    - Dates TBD

- Performance Consulting
  - Performance war rooms: concentrated, highly productive group discussions and analysis
  - MSU reduction services

Like what you see?

- Free z/OS Performance Educational webinars!
  - The titles for our Fall 2022-2023 webinars are as follows:
    - *Key Reports to Evaluate z16 Processor Caches*
    - *Understanding System Recovery Boost's Impact on Performance and Performance Reporting*
    - *WLM Management of DDF Work: What can you do and what has changed?*
    - *Intensity! Understanding the Concepts and Usage of Intensity Measurements*
    - *High, Medium, Low: Understanding how HiperDispatch influences performance in z/OS*
    - *Putting a lid on XCF*
    - *How and why Pivotor is different than other performance management reporters*
    - *Key Reports to Evaluate Usage of Parallel Access Volumes*
    - *Key Reports to Evaluate Coupling Facility CPU Utilization*
    - *Understanding how memory management has evolved in z/OS*
  - Let me know if you want to be on our mailing list for these webinars
- If you want a free cursory review of your environment, let us know!
  - We're always happy to process a day's worth of data and show you the results
  - See also: http://pivotor.com/cursoryReview.html

Like what you see?

- The z/OS Performance Graphs you see here come from Pivotor™

- If you don't see them in your performance reporting tool, or you just want a free cursory performance review of your environment, let us know!
  - We're always happy to process a day's worth of data and show you the results
  - See also: http://pivotor.com/cursoryReview.html

- We also have a free Pivotor offering available as well
  - 1 System, SMF 70-72 only, 7 Day retention
  - That still encompasses over 100 reports!

Why do we care about processor cache measurements and usage of the caching hierarchy?

Key Influences of Processor Performance and Capacity

- Question:

  What are the key influences that result in variations of a particular processor's delivered capacity relative to a customer's environment and workload?

- Answer: As Gary King of IBM would say… there are three key influences:
  - Instruction complexity of one processor family to another
  - Path length of the code executed by customer applications and transactions
  - Usage of the Memory Hierarchy
- A machine's capacity will vary based on each of these three factors

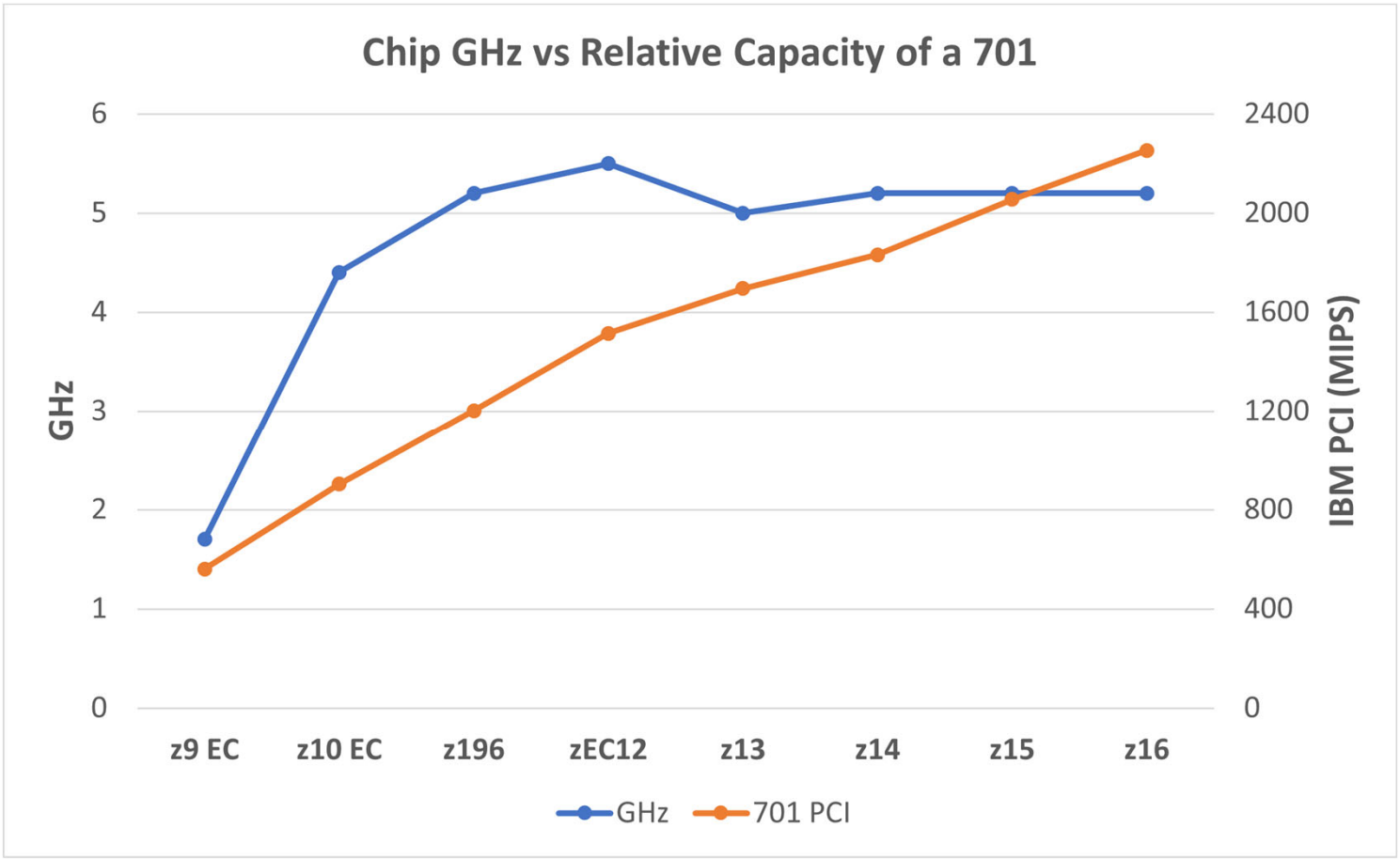
Key Influence - Instruction Complexity

- Notice the PCI column showing the PCIs of the 701 series of each processor
  - PCI : Processor Capacity Index

    Term used by IBM instead of MIPS

							Max per first book-drawer				Processor Cache					
											Core-level				Chip	Book-dwr
zGen	Name	Year	Mach Type	GHz	701 PCI	701 MSUs	Memory	CPs	PU Chips	Cores/chip	L1-Data	L1-Instr	L2-Data	L2-Instr	L3/chip	L4/bk-dwr
z9	z9 EC	2005	2094	1.7	560	81	128G	8	8	2	256K	256K	n/a	n/a	n/a	40M
z10	z10 EC	2008	2097	4.4	902	115	384G	12	5	4	128K	64K	3M		n/a	48M
z11	z196	2010	2817	5.2	1202	150	704G	15	6	4	128K	64K	1.5M		24M	192M
z12	zEC12	2012	2827	5.5	1514	188	704G	20	6	6	96K	64K	1M	1M	48M	348M
z13	z13	2015	2964	5	1695	210	2464G	30	6	8	128K	96K	2M	2M	64M	960M
z14	z14	2017	3906	5.2	1832	227	8000G	33	6	10	128K	128K	4M	2M	128M	672M
z15	z15	2019	8561	5.2	2055	253	8000G	34	4	12	128K	128K	4M	4M	256M	960M
z16	z16	2022	3931	5.2	2253	278	9984G	39	4x2	8	128K	128K	up to 32M		up to 256M	up to 2G

Key Influence - Instruction Complexity



Key Influence – Path Length

- Path length of the code executed by customer applications and transactions
  - This relates to code executed by applications / jobs / transactions / etc.
  - Instruction count
- The actual path lengths executed by a workload will vary
  - From customer to customer, and from IBM synthetic workloads versus customer
  - From one customer's application environment versus another application environment of that same customer
    - Example: CICS / DB2 application versus a WAS / DB2 application
- Is sensitive to the configuration due to MP effects
  - Higher n-ways or difference in configuration may increase path lengths execute (which in turn influences the processor capacity relative to LSPRs)
    - Example: May have more locking in a higher MP environment, or queues may be longer, etc.
- But when move from one processor to another this generally does not change much for a specific customer
  - Whether the move is from one processor family to another
  - Or from one process in the same family to another

Key Influence – Memory Hierarchy

- Usage of the Memory Hierarchy
  - Heavily influenced by key factors result potentially wide variations in realized capacity
  - From one processor family to another there are many design alternatives
    - Levels of cache, scope of cache, latency, etc.
  - Configuration will influence usage of the memory hierarchy
    - LPAR configuration, competition between LPARs, options such as HiperDispatch, etc.
  - Exploitation by workloads will influence usage of the memory hierarchy
    - Transaction intensity, memory intensity, I/O intensity, application mixtures, competition of resource by applications, etc.
  - z/OS performance management and options
    - WLM management of resources, affinity nodes, IEAOPTxx opts, heap sizes, initiators, etc.
- Final result is that usage of memory hierarchy heavily influences a processor's delivered capacity and performance.
  - Workload performance sensitive to how deep into the memory hierarchy the processor must go to retrieve instructions and data
- So, for processor sizing, LSPRs have started focusing on this

Case Study CEC LSPRs: z14 vs z16

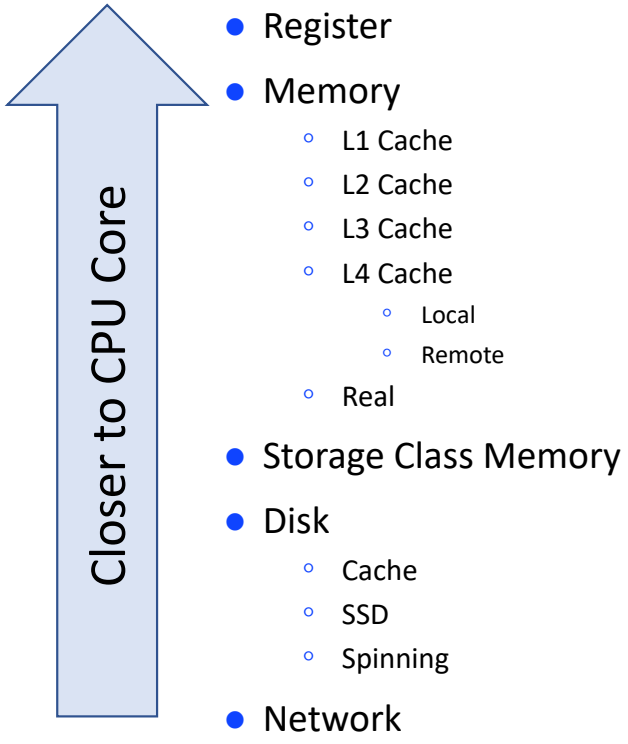
	Processor	#CP	PCI**	MSU***	Low*	Average*	High*
z14	**3906-609**	9	8142	997	15.99	14.55	12.79
	Processor	**#CP**	**PCI****	**MSU*****	**Low***	**Average***	**High***
z16	**3931-606**	6	8006	980	14.92	14.3	13.01

L1MP	RNI	Workload Hint
<3%	>= 0.75	AVERAGE
	< 0.75	LOW
3% to 6%	>1.0	HIGH
	0.6 to 1.0	AVERAGE
	< 0.6	LOW
>6%	>=0.75	HIGH
	< 0.75	AVERAGE

Introducing the Processor Caches of IBM's zArchitecture Processors

Modern Performance Optimization

- Distance matters!
- Keep data close to not just the processor, but close to the instruction units on the processor
  - I.E. L1 cache hits very important
  - If the data isn't in L1, hopefully it's in L2, L3 or L4
- Hardware Instrumentation Services (HIS) records processor efficiency metrics in SMF 113 records
  - Be sure to record these
- SMF 99.14 records record mapping of logical to physical cores
  - Of particular interest for multi-book machines to make sure LPARs aren't crossing books



z15 Cache Summary

z15 CPU Core

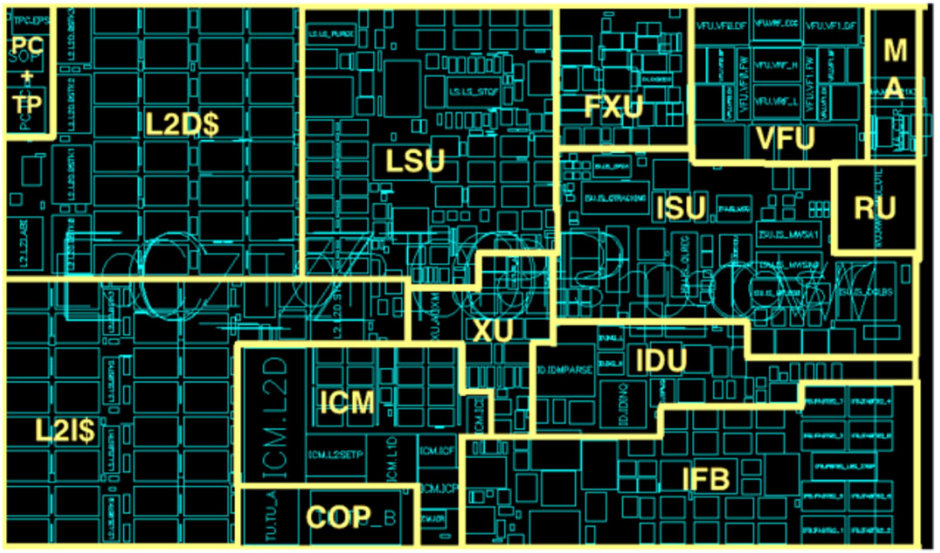
- All processor types (GCP, zIIP, ICF, IFL, SAP, IFP) use the exact same physical core
  - Microcode limits what the individual logical CPs can do
- Note how much area is given to L2 cache
  - 4MB instr, 4MB data
- L1 is in LSU

- 5.2 GHz (0.192ns cycle time)
  - ~5,200,000 cycles in 1ms
  - ~ 2 inches: light speed

ICM: Instr Cache/Merge
IDU: Instr Decode
IFB: Instr fetch/branch predict
ISU: Instr Sequence

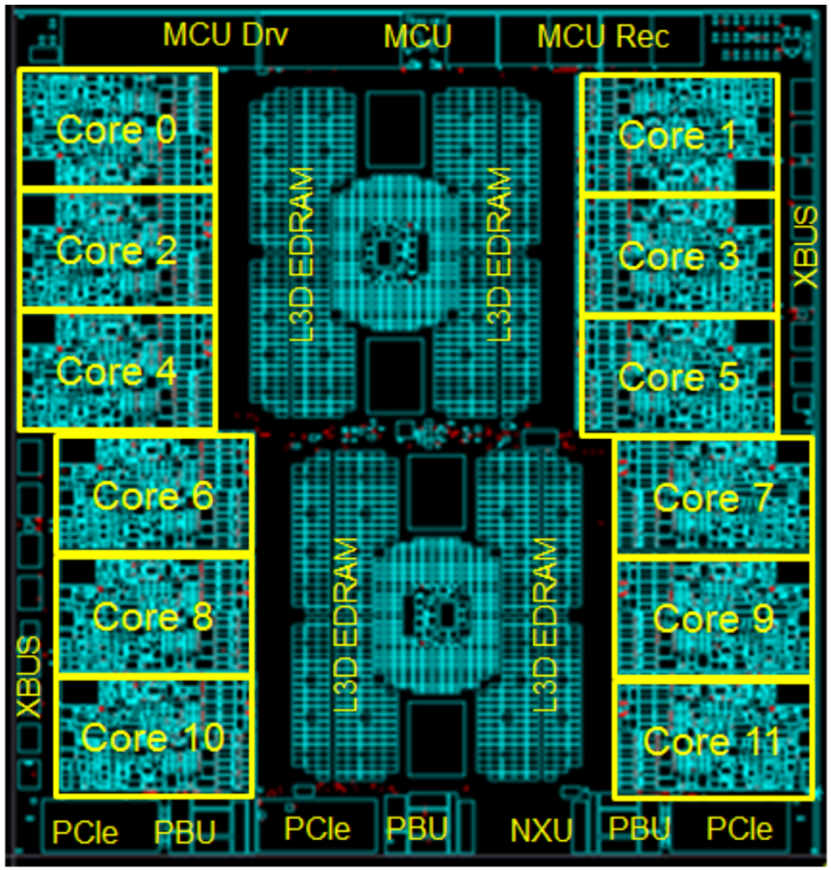
LSU: Load/Store
XU: TLB/DAT
FXU: Fixed point
VFU: Vector / FP
MA: Modulo Arithmetic for Eliptic Curve Crypto

RU: Recovery Unit
COP: CoProc
PC+TP: HIS / error collection, trap



z15 Processor Unit (PU) Chip

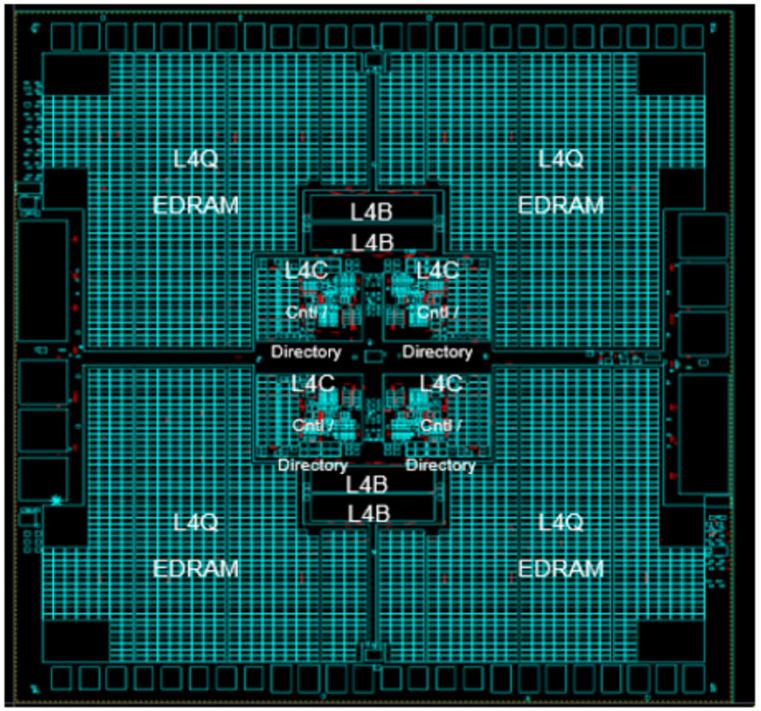
- This is one z15 PU (Processor Unit) Chip
  - About 1” square (25.3mmx27.5mm)
  - 9.2B transistors
- 4 chips per drawer
- 12 cores (9, 10, or 11 “active”) per chip
  - 41 active cores per drawer < Max190
  - 43 active cores per drawer Max190
  - Wafer yields improved by utilizing chips that have some cores disabled
- Notice amount of chip area for L3 cache
  - Note cores rotated to orient L2 near L3
  - Distance matters!
- Note NXU: Nest Acceleration Unit



© IBM

z15 System Controller (SC) Chip

- This is one z14 SC (System Controller) Chip
  - One chip per drawer
- Provides L4 cache
  - 960 MB of cache per SC chip
- Manages communications between PU chips(X-Bus) and drawers (A-Bus)



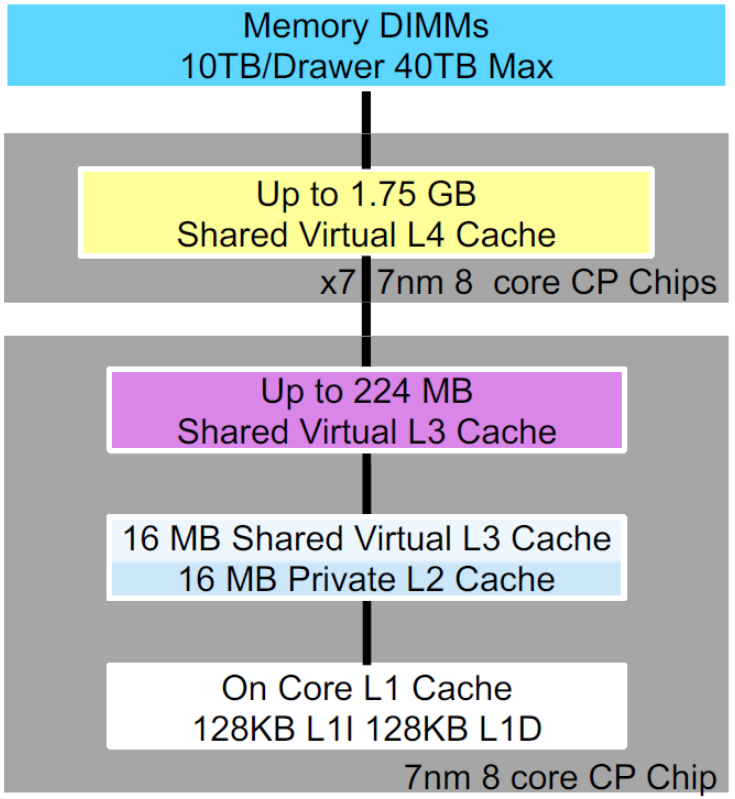
© IBM

z16 Cache Summary

Content for next few z16 slides are from various IBM presentations

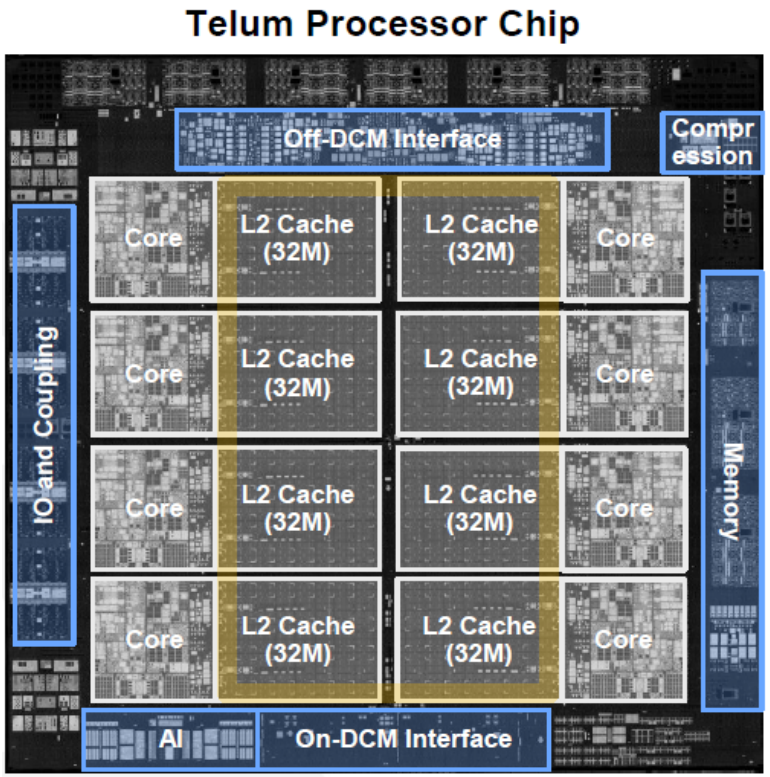
z16 Virtual Caches (slide source: IBM)

- What's different from z15
  - There is no L3 physical cache present on the cores
    - There is a new L1 Shadow Cache that will help manage syncing lines with L2
  - There is no SC chip or physical L4 Cache
    - All CPs L2 are interconnected via buses
- How Virtual Caches work
  - L2 Caches of unused cores or underutilized cores will be converted to be used as virtual caches
    - If the core becomes actives the cache will be returned
  - Virtual cache on the same CP will be seen as additional virtual L3 cache to the core
  - Virtual Cache on a different CP on the same drawer will be seen as L4 Cache



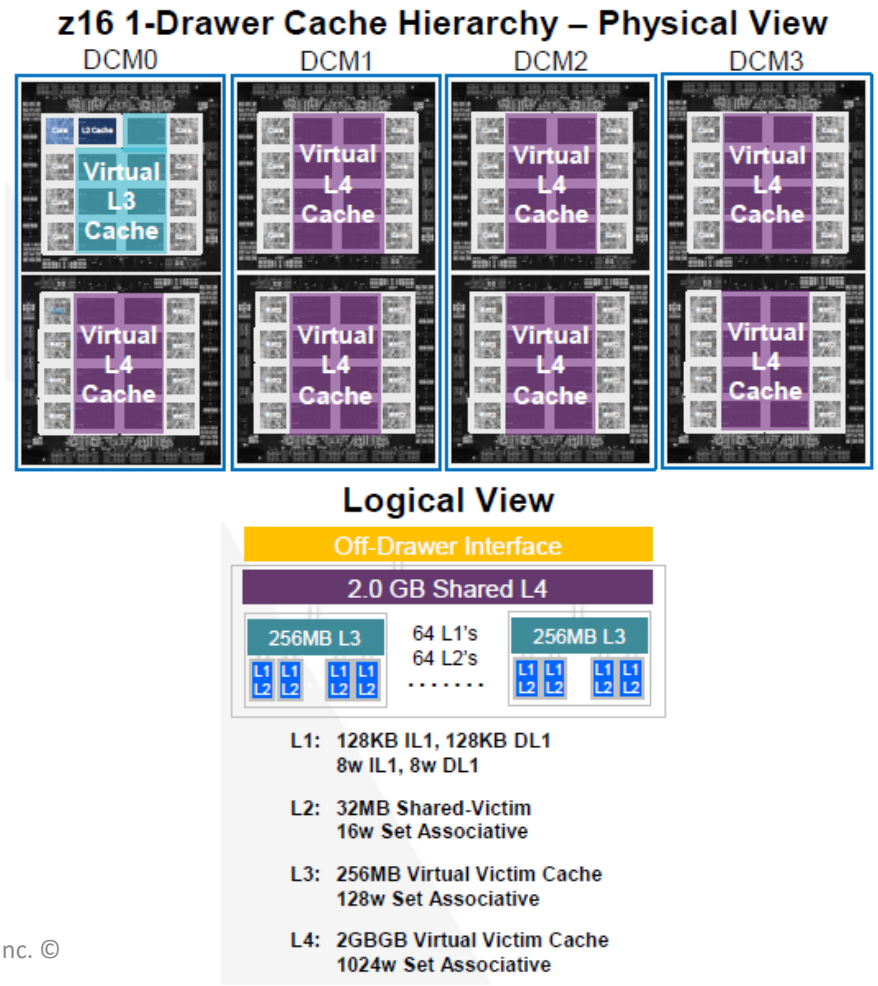
Z16 – Telum Processor Chip (slide source: IBM)

- Samsung 7nm FinFET Technology
  - 530mm 2 chip (25% smaller than
  - Space for future system growth
- 8 Core Processor Chip
  - 4th Generation SMT Core design
  - On Chip Accelerators: AI, Compression, Coupling, Sort
  - Gen 4 PCIe Interface
  - Memory Interface (up to 2TB per chip, encryption +
- Core L1 Cache
  - Private 128KB L1 I and 128KB L1 D
- Core / Chip L2 Cache
  - 32MB Unified L2 cache shared by I & D (4x capacity, 19
  - 4 independent pipelines for fetch/store traffic (320GB/
  - Precise tracking of L1 content (reduces MP
- Scalable high speed, low latency ring interconnect (320GB/



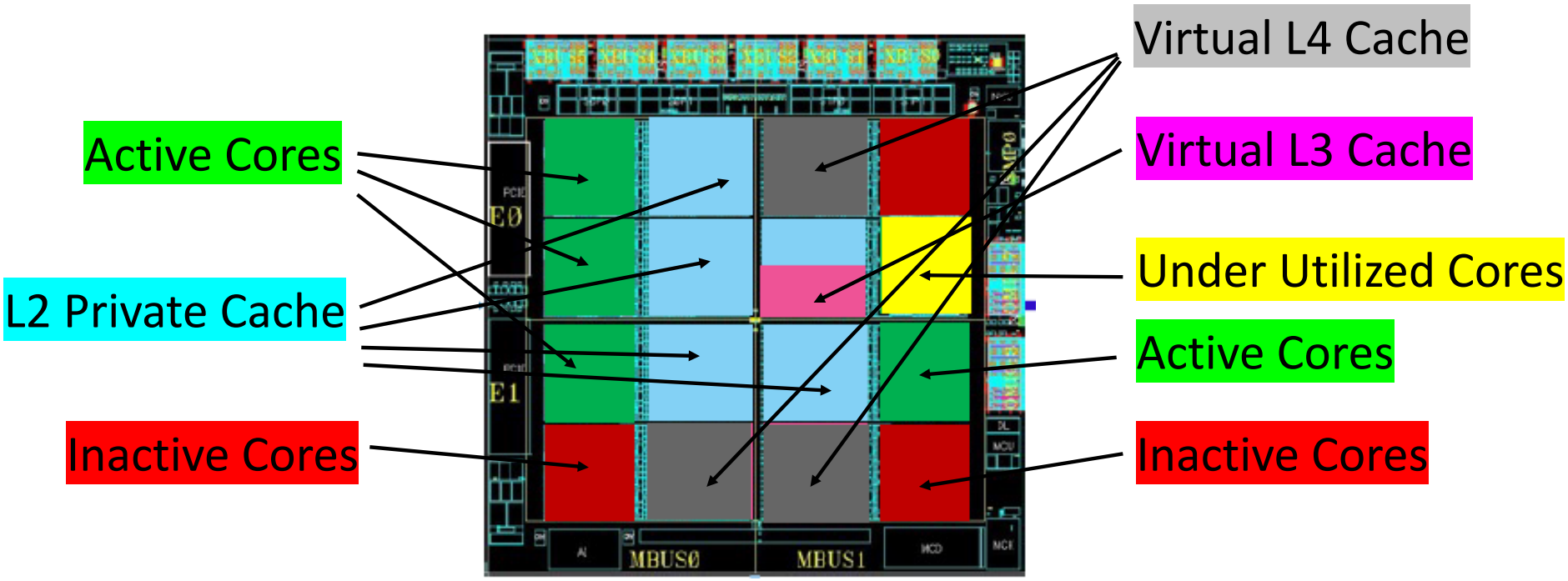
z16 – Virtual L3 and L4 Caches (slide source: IBM)

- Virtual Cache Layers built from L2 Cache blocks
  - Mirrors physical hierarchy of prior designs
- Base Virtual Cache design behavior
  - Private L2 Cache when processor is active (dynamic, 16MB)
  - Virtual L3 Cache when processor is inactive (victim)
  - Virtual L4 Cache when processor chip is inactive (victim)
- Scales as additional processors are brought online
  - L2 Caches switch from Victim L3 to Private L2 behavior
  - Similar to Cache Inclusivity tax of prior designs
- Logical Hierarchy remains
  - 1.5x more cache per core at vL3, vL4
  - More efficient use of cache array space
  - Overcomes limits of traditional architecture
  - Extendable for future generations

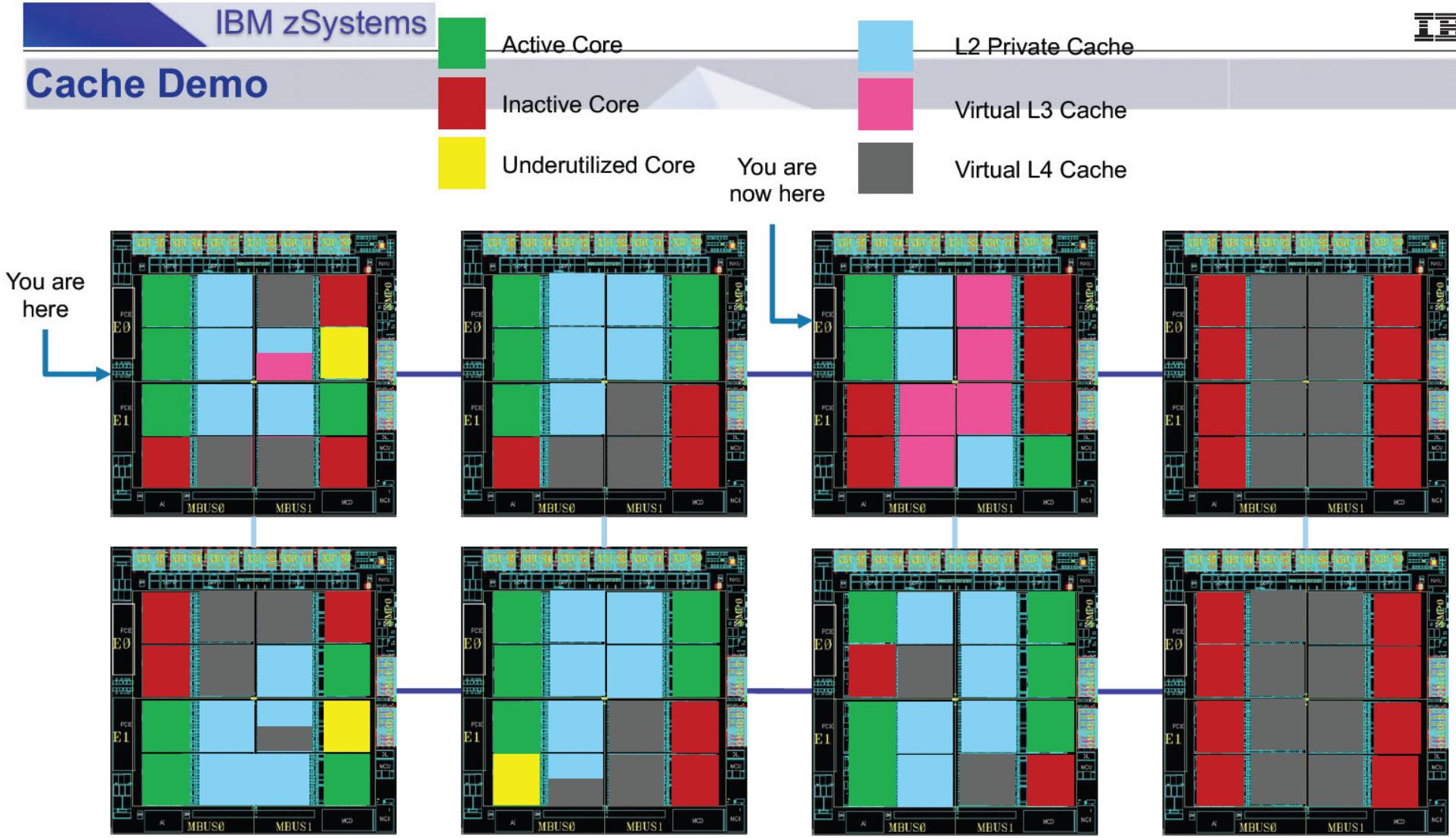


z16 Virtual Cache Provisioning

- One chip example (just to make the point)



Cache Demo

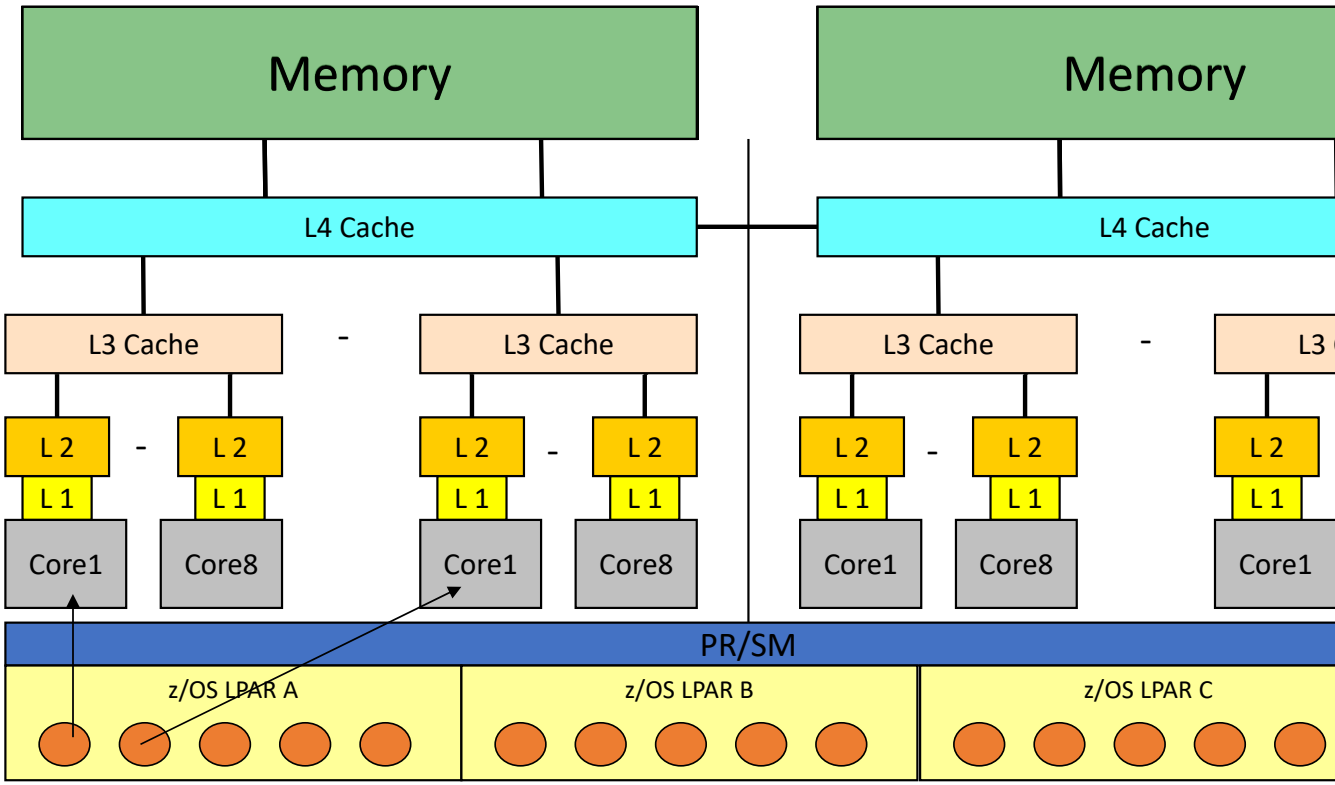


z16 Cache Reporting

Content for next few z16 slides are from various IBM presentations

z16 Processor Logical Cache Hierarchy

- Although virtualized, measurements are calculated for virtual L3 and L4 caches
- Distance Matters!
  - Register
  - Processor Caches
    - L1 Cache
    - L2 Cache
    - L3 Cache
    - L4 Cache
      - Local
      - Remote
    - Real Memory (i.e. Central Storage)
- Hardware Instrumentation Services (HIS) records processor efficiency metrics in SMF 113 records



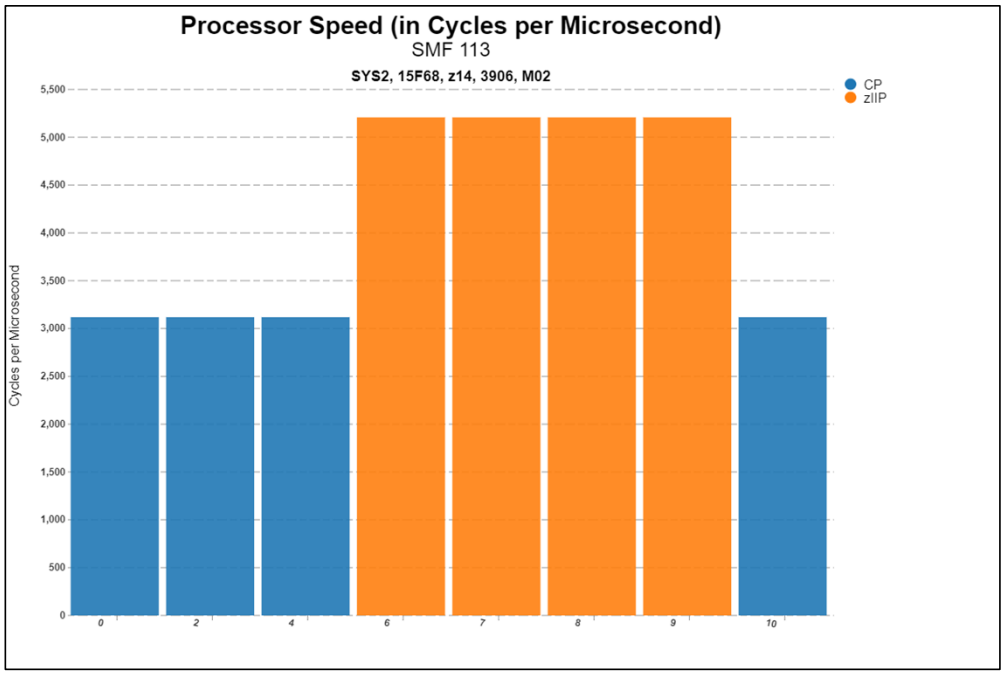
Case Study CEC LSPRs: z14 vs z16

- z14 (3906-609 M02)
- z16 (3931-606 A01)

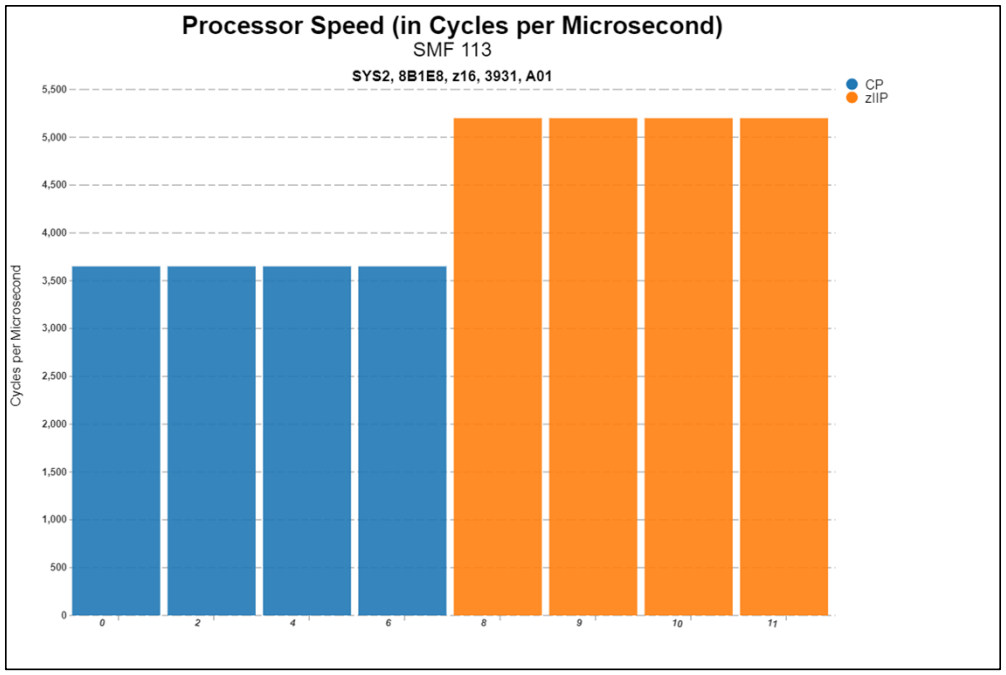
	Processor	#CP	PCI**	MSU***	Low*	Average*	High*
z14	**3906-609**	9	8142	997	15.99	14.55	12.79
	Processor	**#CP**	**PCI****	**MSU*****	**Low***	**Average***	**High***
z16	**3931-606**	6	8006	980	14.92	14.3	13.01

z14 vs z16 SYS2 config

- z14 (3906-609 M02)
  - 9 CPs, 4 zIIPs
  - SYS2: 4 CPs, 4 zIIPs

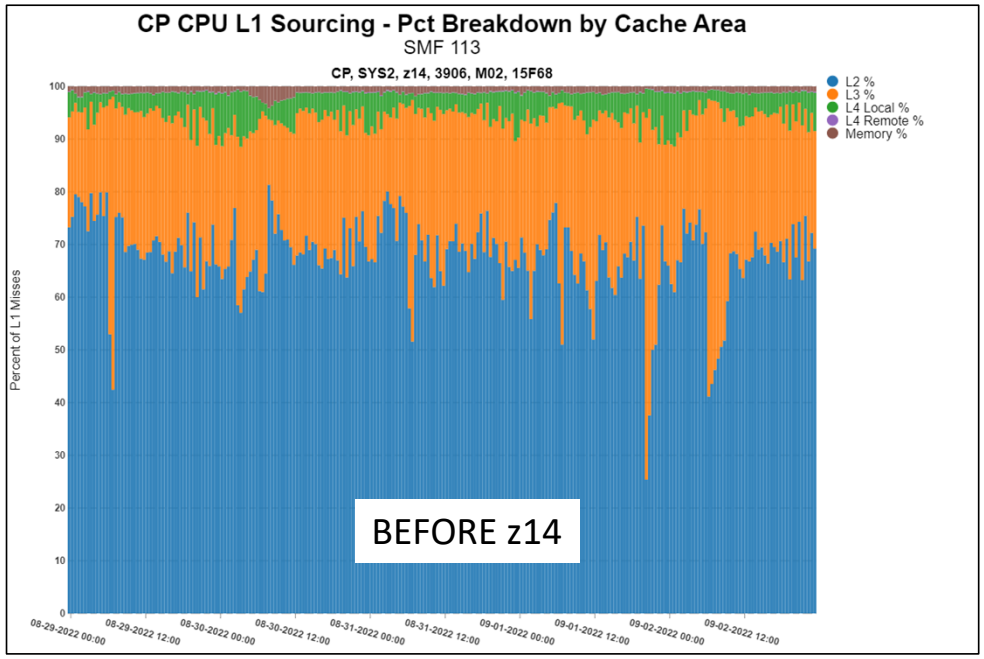


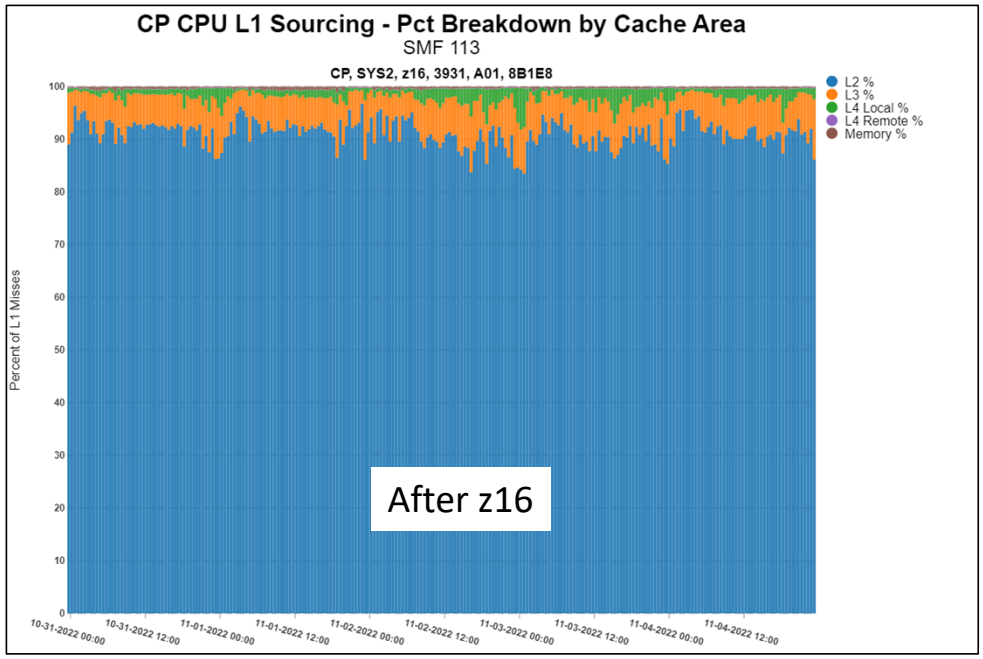
- z16 (3931-606 A01)
  - 6CPs, 4 zIIPs
  - SYS2: 4 CPs, 4 zIIPs



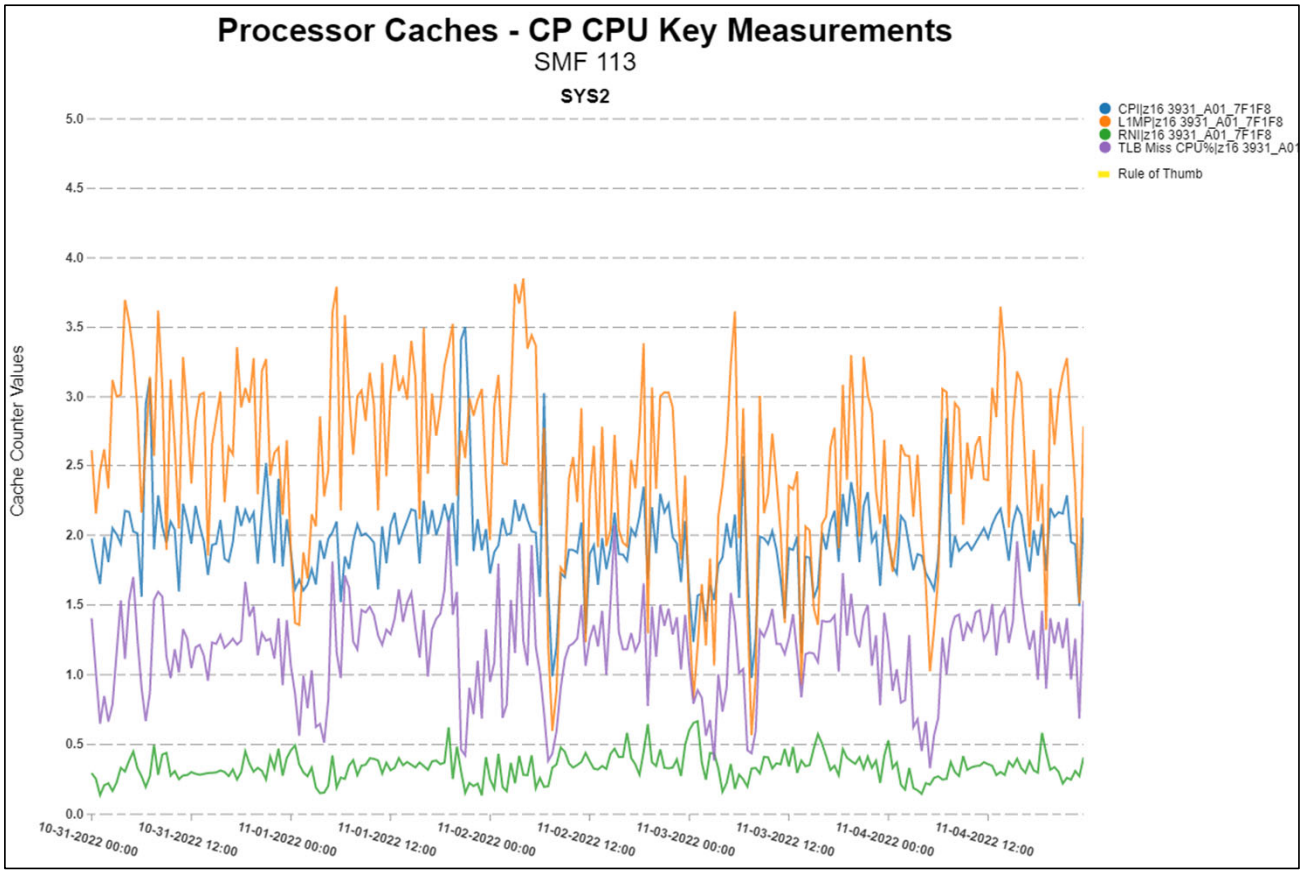
z14 vs z16 – Cache Sourcing

Notice the improved sourcing from L2 since L2 caches are much larger





There are 4 key cache values (Pivotor Example)

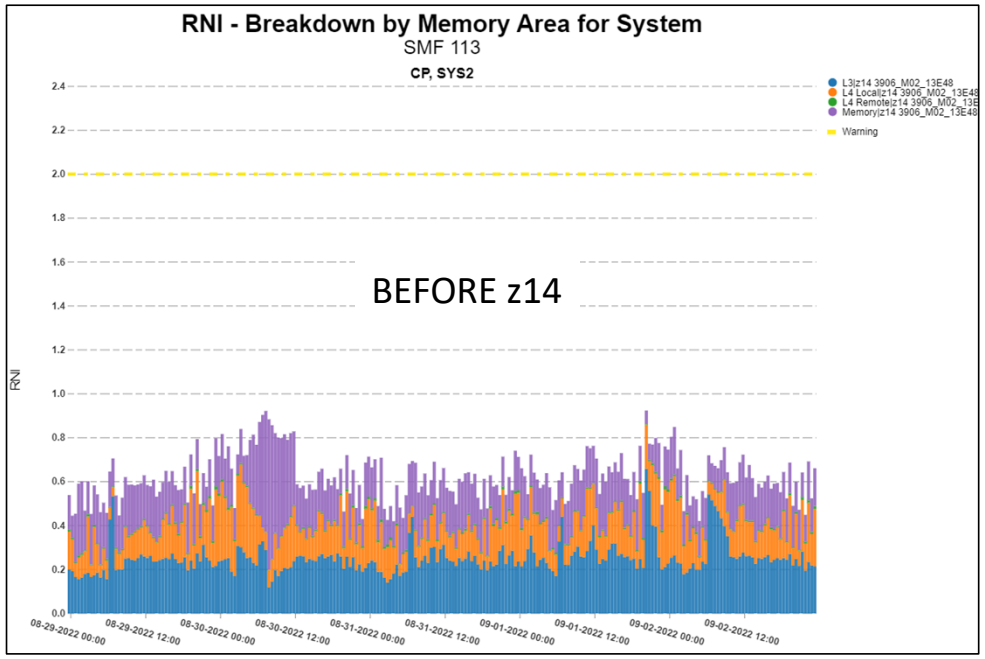


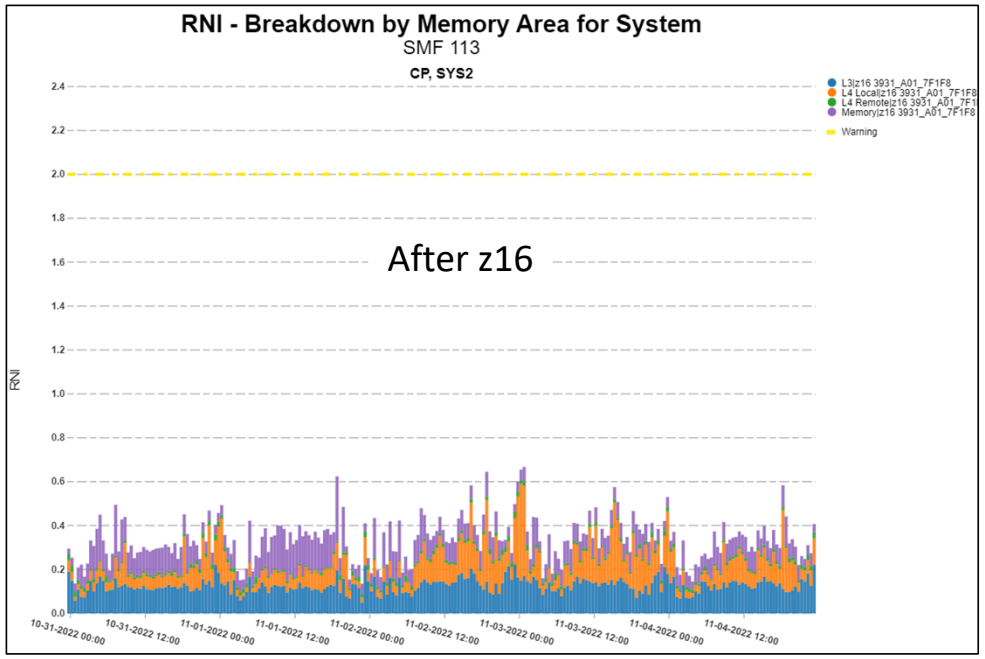
4 key calculated (CP CPU example):

- CPI: Cycles per Instruction
  - Used to gauge processor contention, as well as instruction mix consistency
- L1MP: L1 misses per 100 instructions
  - Think of this as a 'miss percentage'
- RNI: Relative Next Intensity
  - Workload 'signature' that gauges the pressures being put on the upper caches and memory
- TLB Miss CPU Percentage
  - Total percent of the CPU consumed by the LPAR that goes to dynamic address translation (DAT) due to a translation look-aside buffer miss

RNI – Breakdown by Cache

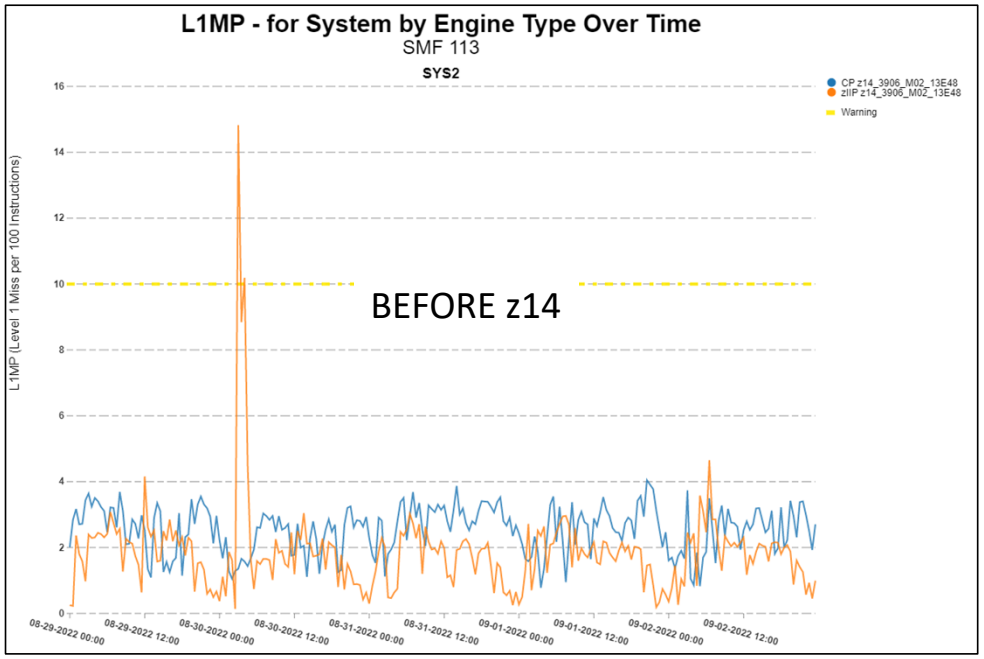
Notice the improved Relative Nest Intensity. Reminder, RNI is not a performance metric to be tuned, but rather a 'signature' of a customer's workloads relative to the LSPRs and machine capacity delivered.

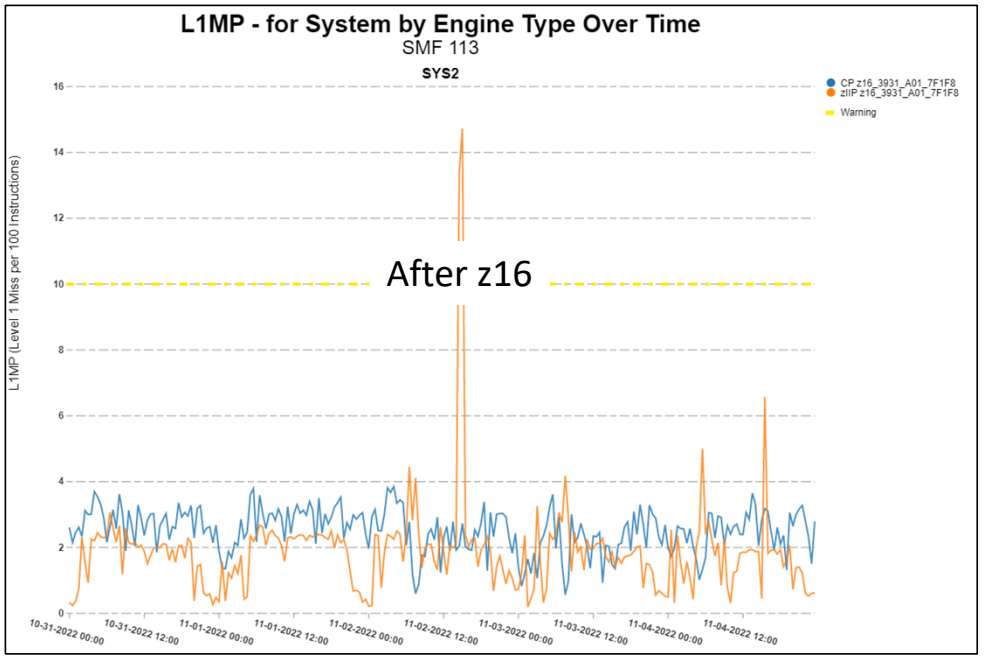




L1MP – z14 vs z16 (L1 Misses per 100 Instr)

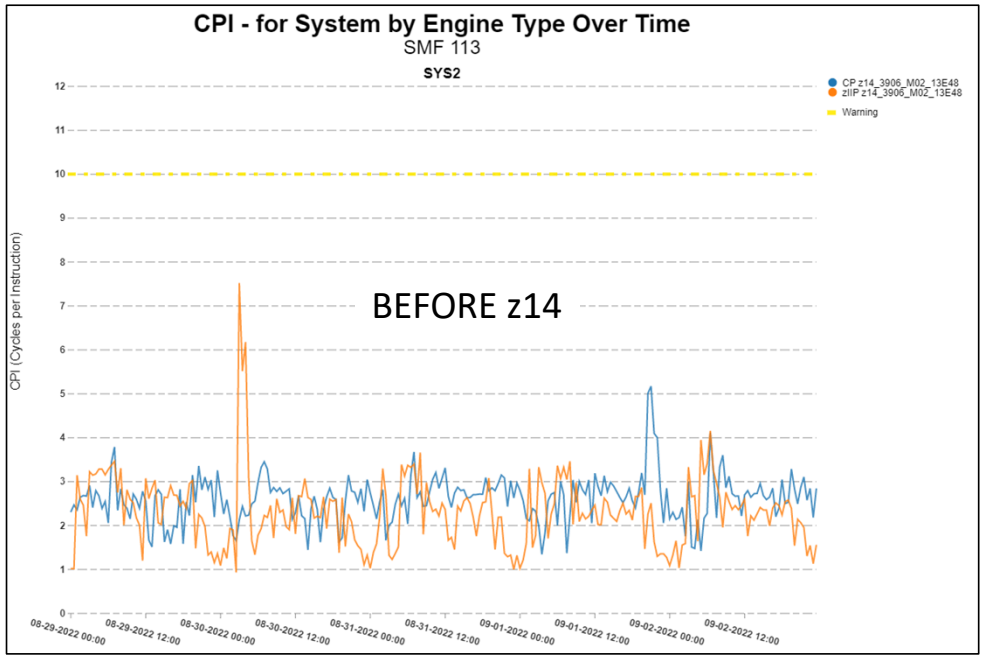
This example shows not that much of a difference for L1 Misses per 100 instructions. Reminder, L1MP is not a performance metric to be tuned, but rather a 'signature' of a customer's workloads relative to the LSPRs.

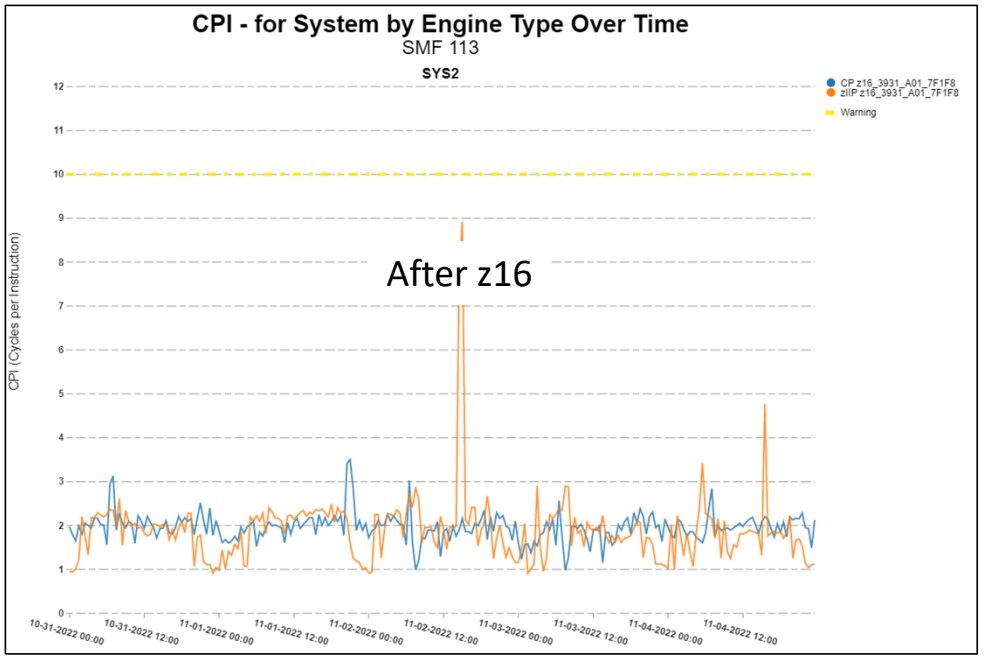




CPI – z14 vs z15 (Cycles per Instruction)

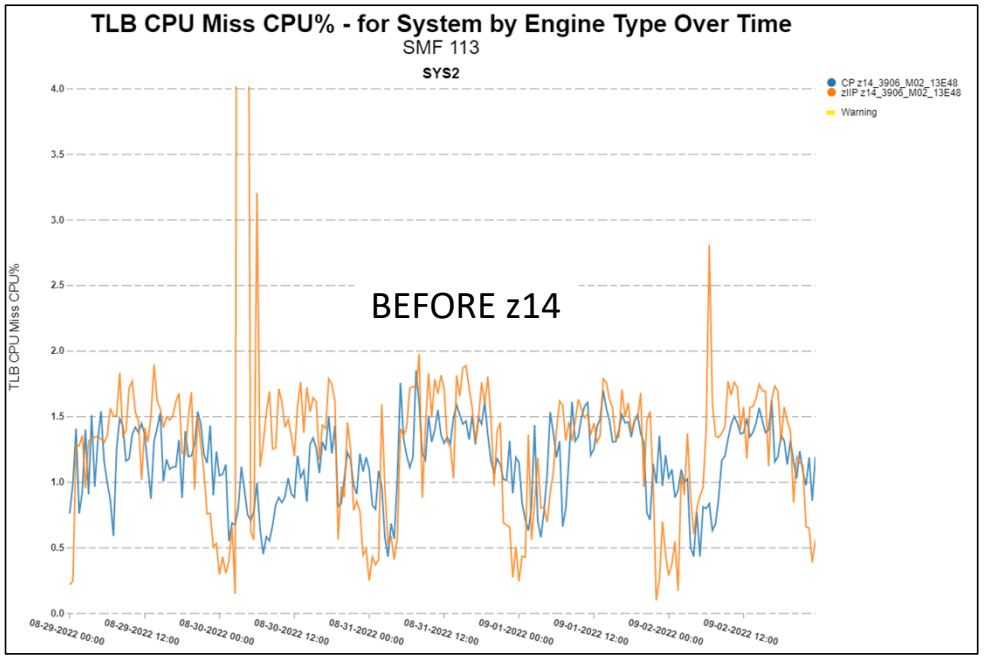
This example shows about a 1 cycle per instruction improvement. Probably mostly due to larger L2 cache

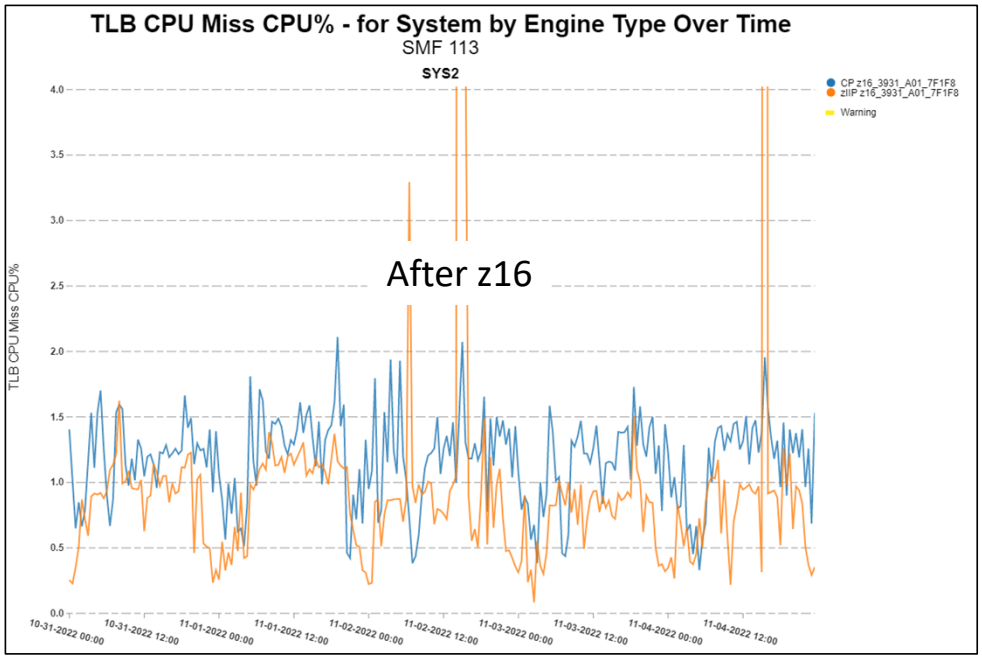




TLB Miss CPU % - z14 vs z15

This example shows TLB CPU Miss % to be about the same at about 2%. This represents pure CPU consumption of Dynamic Address Space (DAT) translation when there is a cache miss in the Translation Lookaside Buffer).





Summary

- The purpose of this presentation was to show some useful caching reports for the z16 processor

- A great deal still needs to be learned

- In addition, the following still needs to be understood, measured, and explained:
  - Current metrics ape measuring the physical cache structure of processors prior to the z16
  - Is it possible to measure the cost of virtualization of the caches
    - Example: cost L4 on chip vs L4 off chip
    - Example: cost of provisioning of the underutilized or inactive L2 to a core's L3 or L4

Comments from Jamie…
and then
Q & A

Questions about content of webinar?

Of maybe general performance questions?